

ENTRADA2

Maarten Wullink

ICANN DNS Symposium

Sep 25, 2024 – Santa Marta, Columbia (Remote)



Introduction

ENTRADA: ENhanced Top-level domain Resilience through Advanced Data Analysis

Enables analysis of large volumes of DNS pcap data, by:

- Converting pcap data to a column-oriented data format.
- Matching and enrichment of DNS queries
- Automated handling of workflow steps

Introduction

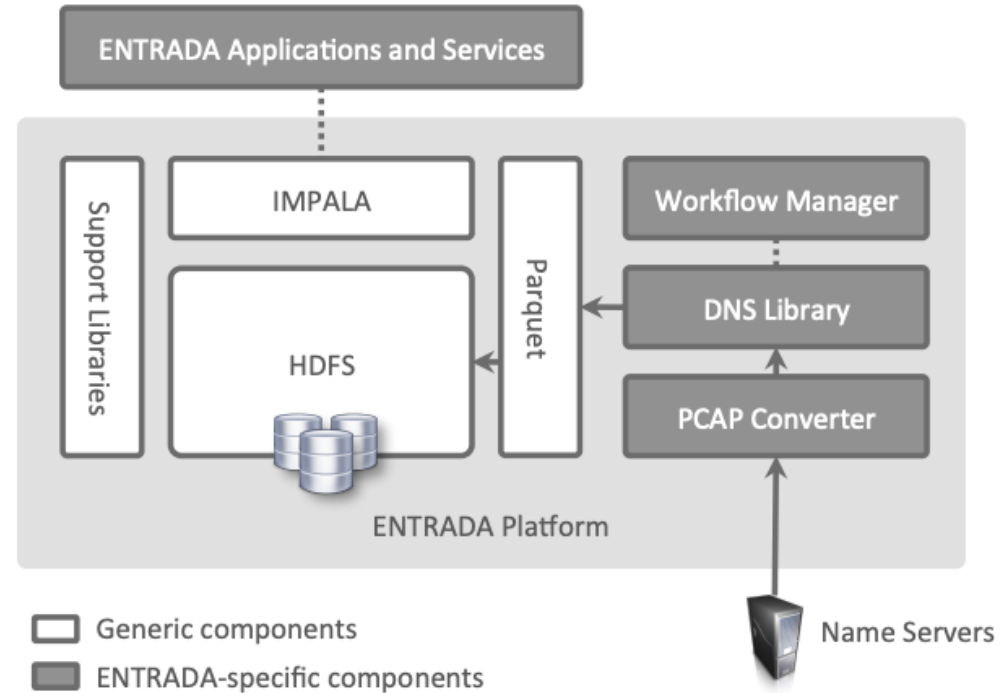
First released as open source 8 years ago, based on:

- Hadoop ecosystem
- Impala + Spark query engines

Mostly used by:

- Domain name (TLD) registry operators
- Researchers

Previous Architecture



Previous Architecture

Based on Hadoop ecosystem, obsoleted by newer technologies

- **HDFS** (Distribute storage) is replaced by s3 compatible Object Storage (AWS, MinIO)
- **YARN** (Process scheduling) is replaced by Kubernetes
- **Kerberos** is replaced by pluggable security providers

ENTRADA2

A new project, based on modern data architecture and cloud native technology

Improvements in area of:

- Scalability (cloud scaling)
- Performance (smaller data footprint)
- Usability (Easier to deploy and maintain than Hadoop)
- Security

ENTRADA2

Major changes:

- No longer depends on Hadoop
- Optimized Parquet data output, 40% size reduction
- Clustering support (Kubernetes, AWS and Docker)
- Use of a lakehouse architecture using open table format (Iceberg)
- Support for any JDBC compatible SQL query engine
- Data model is NOT compatible with original ENTRADA

New Functionality

- Optionally include RDATA of DNS response in output
- Support for decoding Extended rcode
- Support for Extended DNS Errors

ENTRADA2

Multiple deployment modes:

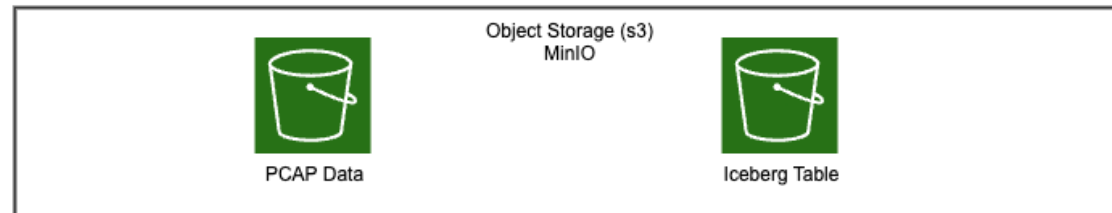
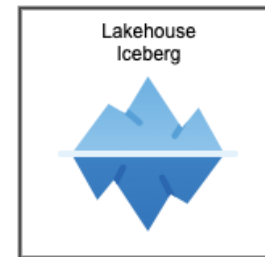
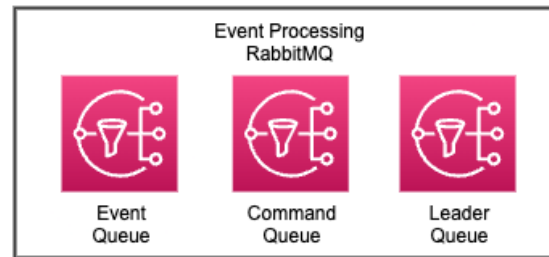
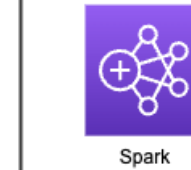
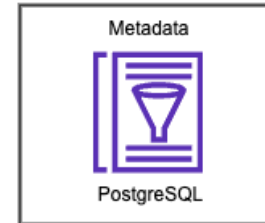
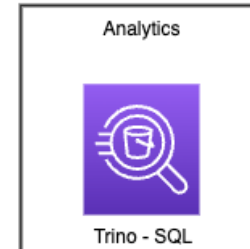
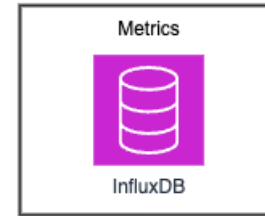
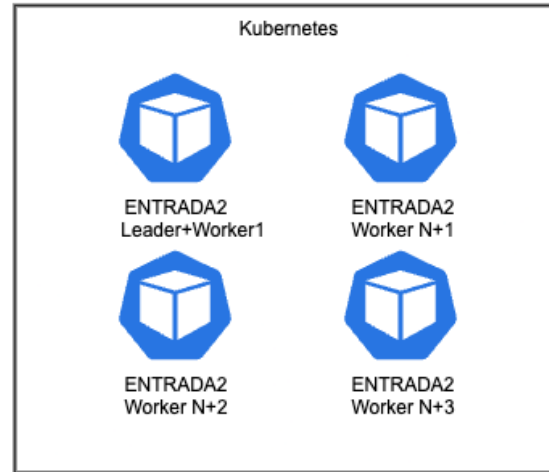
- Kubernetes (on-premise, cloud)
- AWS
- Docker

Easily deploy a cluster using multiple containers

- Auto create required resources (s3, queues, data table)
- Name server site no longer linked to specific container

New Architecture

Kubernetes mode



Lakehouse Architecture



Warehouse-like capabilities on top of a Data Lake provided by Apache Iceberg

- A **Data Lake** is a centralised repository that allows you to store structured, semi-structured, and unstructured data at any scale, using open file formats
- The **Lakehouse** architecture extends the Data Lake by integrating a metadata management layer to provide warehouse-like capabilities

Iceberg



Iceberg Lakehouse features:

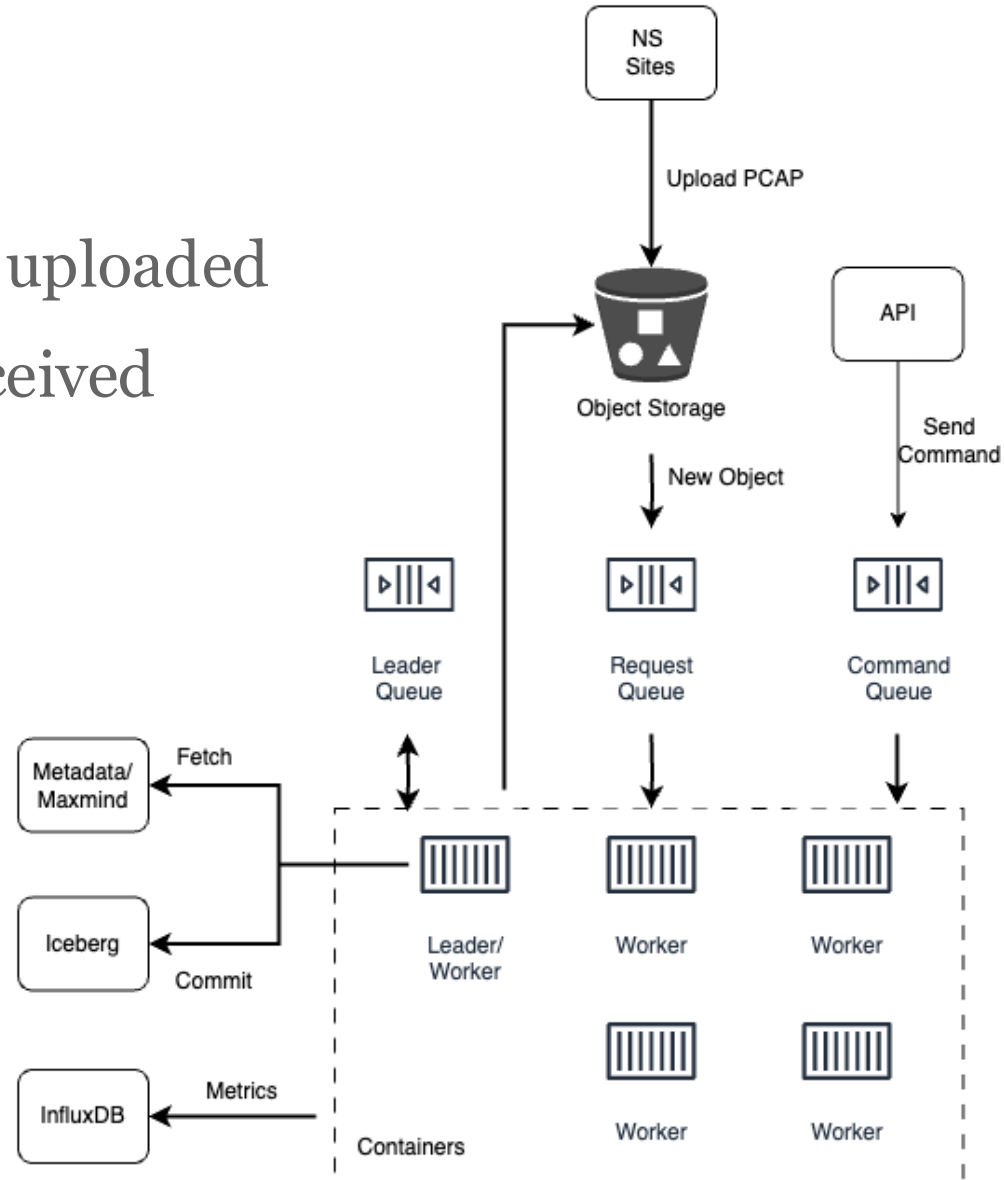
- Schema Evolution
- Hidden Partitioning
- Time Travel
- Transactions
- Data Compaction



Workflow

Event-driven

- New PCAP object uploaded
- API command received



Demo

Run docker-compose.yml to start new 2-node cluster

- 1 Leader/worker
- 1 Worker

and dependencies:

- MinIO (s3)
- Trino (SQL engine)
- InfluxDB (metrics)
- RabbitMQ (messaging)
- PostgreSQL (meta data)

```
docker-compose
entrada-master-1 | 35: edns_ecs_ip_asn: optional string
entrada-master-1 | 36: edns_ecs_ip_asn_org: optional string
entrada-master-1 | 37: edns_ecs_ip_geo_country: optional string
entrada-master-1 | 38: edns_ext_error: optional list<int>
entrada-master-1 | 40: dns_labels: optional int
entrada-master-1 | 41: dns_proc_time: optional int
entrada-master-1 | 42: dns_pub_resolver: optional string
entrada-master-1 | 43: dns_req_len: optional int
entrada-master-1 | 44: dns_res_len: optional int
entrada-master-1 | 45: top_ttl: optional int
entrada-master-1 | 46: server: required string
entrada-master-1 | 47: server_location: required string
entrada-master-1 | 48: dns_rdata: optional list<struct<0: section: required int, 51: type: required int, 52: data: optional string>>
entrada-master-1 | }
entrada-worker-2 | 2024-09-23T10:00:07.668Z INFO 1 --- [entrada2] [ main ] o.s.a.r.c.CachingConnectionFactory : Attempting to connect to: [rabbitmq:5672]
rabbitmq-1 | 2024-09-23 10:00:07.717440+00:00 [info] <0.754.0>- accepting AMQP connection <0.754.0> (172.18.0.9:46678 -> 172.18.0.2:5672)
entrada-worker-3 | 2024-09-23T10:00:07.757Z INFO 1 --- [entrada2] [ main ] nl.sidn.entrada2.StartupListener : This pod is working correct
rabbitmq-1 | 2024-09-23 10:00:07.913089+00:00 [info] <0.754.0>- connection <0.754.0> (172.18.0.9:46678 -> 172.18.0.2:5672) has a client-provided name: rabbitConnectionFactory#43414b88:0
rabbitmq-1 | 2024-09-23 10:00:07.923758+00:00 [info] <0.754.0>- connection <0.754.0> (172.18.0.9:46678 -> 172.18.0.2:5672) - rabbitConnectionFactory#43414b88:0: user 'admin' authenticated and granted access to vhost '/'
entrada-worker-2 | 2024-09-23T10:00:07.937Z INFO 1 --- [entrada2] [ main ] o.s.a.r.c.CachingConnectionFactory : Created new connection: rabbitConnectionFactory#43414b88:0/SimpleConnection#83c464d [del
egate=amp://admin@172.18.0.2:5672/, localPort=46678]
entrada-worker-2 | 2024-09-23T10:00:07.955Z INFO 1 --- [entrada2] [ main ] o.s.amqp.rabbit.core.RabbitAdmin : Auto-declaring a non-durable, auto-delete, or exclusive Queue (spring.gen-elKUGudQmY834
Cr(4)BA) durable=false, auto-delete=true, exclusive=false. It will be redeclared if the broker stops and is restarted while the connection factory is alive, but all messages will be lost.
entrada-worker-2 | 2024-09-23T10:00:08.112Z INFO 1 --- [entrada2] [ main ] nl.sidn.entrada2.StartupListener : This is NOT the leader, make sure not to be listening to leader queue
entrada-worker-2 | 2024-09-23T10:00:08.112Z INFO 1 --- [entrada2] [ main ] n.s.e.s.messaging.AbstractRabbitQueue : Stopping queue: entrada-leader
entrada-worker-2 | 2024-09-23T10:00:08.113Z INFO 1 --- [entrada2] [ main ] nl.sidn.entrada2.Application : Started Application in 22.023 seconds (process running for 24.091)
entrada-worker-1 | 2024-09-23T10:00:08.264Z INFO 1 --- [entrada2] [ main ] o.s.i.endpoint.EventDrivenConsumer : Adding [logging-channel-adapter:org.springframework.integration.errorLogger] as a subscri
ber to the 'errorChannel' channel.
entrada-worker-1 | 2024-09-23T10:00:08.266Z INFO 1 --- [entrada2] [ main ] o.s.i.channel.PublishSubscribeChannel : Channel 'entrada2.errorChannel' has 1 subscriber(s).
entrada-worker-1 | 2024-09-23T10:00:08.267Z INFO 1 --- [entrada2] [ main ] o.s.i.endpoint.EventDrivenConsumer : started bean 'org.springframework.integration.errorLogger'
entrada-worker-1 | 2024-09-23T10:00:08.286Z INFO 1 --- [entrada2] [ main ] o.s.b.w.embedded.tomcat.TomcatWebServer : Tomcat started on port 8080 (http) with context path '/api/v1'
entrada-worker-1 | 2024-09-23T10:00:08.296Z INFO 1 --- [entrada2] [ main ] o.s.a.r.c.CachingConnectionFactory : Attempting to connect to: [rabbitmq:5672]
rabbitmq-1 | 2024-09-23 10:00:08.346531+00:00 [info] <0.780.0>- accepting AMQP connection <0.780.0> (172.18.0.10:42110 -> 172.18.0.2:5672)
entrada-master-1 | 2024-09-23T10:00:08.437Z INFO 1 --- [entrada2] [ main ] o.s.b.a.e.web.EndpointLinksResolver : Exposing 1 endpoint beneath base path '/actuator'
rabbitmq-1 | 2024-09-23 10:00:08.548333+00:00 [info] <0.780.0>- connection <0.780.0> (172.18.0.10:42110 -> 172.18.0.2:5672) has a client-provided name: rabbitConnectionFactory#432b883:0
rabbitmq-1 | 2024-09-23 10:00:08.560514+00:00 [info] <0.780.0>- connection <0.780.0> (172.18.0.10:42110 -> 172.18.0.2:5672) - rabbitConnectionFactory#432b883:0: user 'admin' authenticated and granted access to vhost '/'
entrada-worker-2 | 2024-09-23T10:00:08.589Z INFO 1 --- [entrada2] [ main ] nl.sidn.entrada2.StartupListener : This pod is working correct
entrada-worker-1 | 2024-09-23T10:00:08.610Z INFO 1 --- [entrada2] [ main ] o.s.a.r.c.CachingConnectionFactory : Created new connection: rabbitConnectionFactory#432b883:0/SimpleConnection#6c31ed81 [del
egate=amp://admin@172.18.0.2:5672/, localPort=42110]
entrada-worker-1 | 2024-09-23T10:00:08.630Z INFO 1 --- [entrada2] [ main ] o.s.amqp.rabbit.core.RabbitAdmin : Auto-declaring a non-durable, auto-delete, or exclusive Queue (spring.gen-tUzDfYzRv6J019
871257g) durable=false, auto-delete=true, exclusive=false. It will be redeclared if the broker stops and is restarted while the connection factory is alive, but all messages will be lost.
entrada-worker-1 | 2024-09-23T10:00:08.802Z INFO 1 --- [entrada2] [ main ] nl.sidn.entrada2.StartupListener : This is NOT the leader, make sure not to be listening to leader queue
entrada-worker-1 | 2024-09-23T10:00:08.802Z INFO 1 --- [entrada2] [ main ] n.s.e.s.messaging.AbstractRabbitQueue : Stopping queue: entrada-leader
entrada-worker-1 | 2024-09-23T10:00:08.804Z INFO 1 --- [entrada2] [ main ] nl.sidn.entrada2.Application : Started Application in 22.45 seconds (process running for 24.667)
```



Demo

Data processing

- Upload new PCAP to s3 bucket
- Monitor s3 event processing
- Check results
 - Data in s3 bucket
 - Metrics in InfluxDB

Analyze data using

- Trino command line
- Dbeaver SQL-client

More details

For more information and getting started details, see:

<https://github.com/SIDN/entrada2>

 SIDN.nl

 @SIDN

 SIDN

Thank You

www.sidnlabs.nl | stats.sidnlabs.nl

